

Amendments to the Specification:

Please replace paragraph 31 beginning on page 11 with the following:

The other component part of the document, referred to here as metadata 210, represents specific properties or characteristics associated with the document. Metadata may be entered by a user when creating a document, or may be entered through default settings in an application program. For example, one might discover metadata by opening a "properties" portion of an opened document, or by some similar method. Thus, metadata 210 are meant to give information that can help characterize the contents in some way. By way of example, the metadata 210 in Figure 2 include metadata concepts (or "segments"), such as an "Author" segment ~~210-212~~, a file "Size" segment 214, and a "Keywords" segment 216. The metadata concepts are like classifications in the sense that they can be more generalized concepts of the more specific, underlying data content 205. Thus, a class is appropriately defined as a set of documents, and a label (e.g., all documents "Authored" by person A).

Please replace paragraph 41 beginning on page 16 with the following:

If the segment does not signal the end of the document 412, then the gatherer processes the segment further to determine if the segment matches one of the class types 420 (e.g., "By AEJ"). If the segment appears to be a valid class type 425, then the gatherer will seek to identify 430 whether there is a known class alias for the apparent class type (e.g., "By=Author=From=Written_By"). If there is no known class alias for the segment so that "By" represents any specific meaning, the gatherer will move to the next segment 450 for processing. If on the other hand, there is 445 an alias identifying the class type (e.g., "By" is recognized as

an “Author” metadata concept), then the gatherer can seek to identify 440 whether a normalized value is found. If a normalized value is found, the gatherer then associates 460 the segment with the appropriate class by mapping the segment value to a preferred class label, and looks to see if any metadata associated with the class type are also aliases (i.e., “normalized” for a standard value – “AEJ = ‘Aaron E. Jones’”) so that the class type property may be called for the normalized metadata (or alias for the normalized metadata) and vice versa. Having associated 460 the segment with any available classes and/or corresponding metadata, the gatherer adds the segment to a searchable index (e.g., 360) by association with the normalized value, and then proceeds to get the next segment 450. If on the other hand, no normalized value is found, the gatherer may not add a new segment and may proceed to get the next segment 450.

Please replace paragraph 42 beginning on page 17 with the following:

If the next segment 450 resembles an end of document designator ~~440~~ 415, the gatherer determines if there is a next document 470 to filter into segments. For example, the gatherer may detect that there is an embedded URI for a next document (e.g., “crawling”), or the gatherer may detect a next document as input received in queue from another input method (e.g., a file list). If the gatherer detects a next document, 495, then the gatherer proceeds to filter the next document into segments, and continues the previously-described processing path. Alternatively, if the gatherer detects no new document, the gatherer ceases processing documents 480.

Please replace paragraph 47 beginning on page 19 with the following:

Act 540 may be performed by associating the received document with the document alias so that, upon request for the document datum through a search engine, the received document is returned to the requester by association of the document datum with the alias. Thus, for example, if a person enters the term "Nathan Smith" into a search engine input box and then executes the search, the search engine will look for any occurrence of the term "Nathan Smith." The search engine will also look for all documents containing the associated aliases (e.g., terms "NDS," and "NAS," from contact entry ~~242~~ 252 of Figure 2), since the term "Nathan Smith," and each of the received aliases have been normalized. 242 is a contact entry associated with the name Aaron Jones, as shown in Figure 2.